Math-Net.Ru

Общероссийский математический портал

И. А. Лёзин, Разложение двумерных плотностей вероятностей в ортогональных базисах, *Вестн. Сам. гос. техн. ун-та. Сер. Физ.-мат. науки*, 2009, выпуск 1(), 169–174

DOI: <https://doi.org/10.14498/vsgtu654>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением

<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.237.206.68

26 апреля 2023 г., 09:54:15

УДК 517.587:519.216

РАЗЛОЖЕНИЕ ДВУМЕРНЫХ ПЛОТНОСТЕЙ ВЕРОЯТНОСТЕЙ
В ОРТОГОНАЛЬНЫХ БАЗИСАХ

И. А. Лёзин

Самарский государственный аэрокосмический университет им. ак. С. П. Королёва,
443086, г. Самара, Московское ш., 34.

E-mail: chuchyck@yandex.ru

Исследован подход к аппроксимации двумерных плотностей вероятностей на прямоугольных областях с помощью семейств двумерных ортогональных функций, полученных как произведение одномерных ортогональных функций. Приводятся формулы расчета коэффициентов аппроксимации в различных базисах при первичном приближении входных данных в виде сплайна, а также величина методической погрешности при выбранных значениях коэффициентов переноса и масштаба.

Ключевые слова: аппроксимация, ортогональный базис, двумерная функция, коэффициент разложения, плотность вероятностей, двумерный сплайн.

При работе с большими выборками однородных данных возникает необходимость получения функциональных вероятностных характеристик, например, таких как плотность вероятностей, которые используются при анализе свойств имеющихся данных. Непрерывная функциональная модель, построенная по эмпирической выборке, позволяет перейти к аналитическим выкладкам при решении ряда задач. Например, в машиностроении, имея выражение плотности вероятностей для распределения значений дефектов отдельных деталей, можно перейти к совместной плотности узла, состоящего из нескольких деталей, и заранее оценить вероятность и величину отклонений от нормы при сборке.

Рассмотрим случай с выборкой, представленной набором двумерных данных, например, координат случайных точек на плоскости. Первичное приближение плотности вероятностей представим в виде сплайна степени p , построенного на основе двумерной гистограммы:

$$s^{(p)}(x, y) = s_{ij}^{(p)}(x, y), \quad x_i \leq x < x_{i+1}, \quad y_j \leq y < y_{j+1}, \quad (1)$$

$$i = 0, 1, \dots, N; \quad j = 0, 1, \dots, M,$$

где

$$s_{ij}^{(p)}(x, y) = \sum_{l=0}^p \sum_{m=0}^p a_{ij}^{(lm)} x^l y^m. \quad (2)$$

Разложим плотность вероятностей в ряд вида

$$\hat{f}(x, y) = \sum_{n=0}^N \sum_{m=0}^{N-n} \beta_{nm} F_{nm}(x, y), \quad (3)$$

Лёзин Илья Александрович — ассистент кафедры информационных систем и технологий.

где $F_{nm}(x, y)$ — семейство функций [1, 2], ортонормированных на некоторой прямоугольной области G с весовой функцией $\mu(x, y)$.

Имея выражение (1) в качестве первичного приближения, коэффициенты разложения функции в ряд вида (3) можно оценить по формуле

$$\hat{\beta}_{nm} = \iint_G s^{(p)}(x, y) F_{nm}(x, y) \mu(x, y) dx dy. \quad (4)$$

Однако в таком виде формула справедлива только в том случае, когда область G совпадает с областью определения сплайна D . Введём понятие функций, ортонормированных на произвольной прямоугольной области:

$$\begin{aligned} \hat{F}_{nm}(x, y) &= \frac{1}{\sqrt{\alpha_x \alpha_y}} F_{nm}\left(\frac{x - \beta_x}{\alpha_x}, \frac{y - \beta_y}{\alpha_y}\right) = \frac{1}{\sqrt{\alpha_x}} P_n\left(\frac{x - \beta_x}{\alpha_x}\right) \frac{1}{\sqrt{\alpha_y}} Q_m\left(\frac{y - \beta_y}{\alpha_y}\right), \\ \hat{\mu}(x, y) &= \mu\left(\frac{x - \beta_x}{\alpha_x}, \frac{y - \beta_y}{\alpha_y}\right) = \mu_x\left(\frac{x - \beta_x}{\alpha_x}\right) \mu_y\left(\frac{y - \beta_y}{\alpha_y}\right). \end{aligned} \quad (5)$$

Здесь $P_n(x)$ и $Q_m(y)$ — семейства одномерных ортонормированных функций с весовыми функциями $\mu_x(x)$ и $\mu_y(y)$; α_x и α_y — коэффициенты масштаба; β_x и β_y — коэффициенты переноса.

В рамках работы рассматриваются полиномы Лежандра и Чебышева первого и второго рода, определенные на интервале $[-1; 1]$, функции Лагерра, заданные на интервале $[0; \infty)$, и функции Эрмита, ортогональные на бесконечном интервале. Общий вид рассматриваемых одномерных функций указан в табл. 1.

Таблица 1

Виды одномерных функций	
Базис	Общий вид k -того элемента
Лежандра, $\mu(x) = 1$	$P_k(x) = \sum_{s=0}^{[k/2]} a_s x^{k-2s},$ $a_s = (-1)^s \frac{\sqrt{2k+1}}{\sqrt{2}} \frac{(2k-2s)!}{2ks!(k-s)!(k-2s)!}$
Чебышева первого рода, $\mu(x) = \frac{1}{\sqrt{1-x^2}}$	$T_0(x) = \frac{1}{\sqrt{\pi}}, T_k(x) = \sum_{s=0}^{[k/2]} a_s x^{k-2s},$ $a_s = (-1)^s \sqrt{\frac{2}{\pi}} 2^{k-2s-1} \frac{k(k-s-1)!}{s!(k-2s)!}$
Чебышева второго рода, $\mu(x) = \sqrt{1-x^2}$	$U_0(x) = \frac{2}{\sqrt{\pi}}, U_k(x) = \sum_{s=0}^{[k/2]} a_s x^{k-2s},$ $a_s = (-1)^s \sqrt{\frac{2}{\pi}} 2^{k-2s} \frac{(k-s)(k-s-1)!}{s!(k-2s)!}$
Лагерра, $\mu(x) = 1$	$L_0(x) = e^{-x/2}, L_k(x) = \sum_{s=0}^{[k/2]} a_s x^s e^{-x/2},$ $a_s = (-1)^s \frac{k!}{(s!)^2 (k-s)!}$
Эрмита, $\mu(x) = 1$	$H_k(x) = \sum_{s=0}^{[k/2]} a_s x^{k-2s} e^{-x^2/2},$ $a_s = (-1)^s \sqrt{\frac{k!}{2^k \sqrt{\pi}}} \frac{2^{k-2s}}{s!(k-2s)!}$

Необходимо найти такие значения коэффициентов переноса β_x и β_y , коэффициентов масштаба α_x и α_y , чтобы обеспечить ортонормальность семейства функций (5) и минимум погрешности аппроксимации:

$$\begin{aligned}\Delta_{\text{мет}}^{(N)} &= \iint_D \left(s^{(p)}(x, y) - \hat{f}(x, y) \right)^2 \hat{\mu}(x, y) dx dy = \\ &= \iint_D \left(s^{(p)}(x, y) \right)^2 \hat{\mu}(x, y) dx dy - \sum_{n=0}^N \sum_{m=0}^{N-n} \hat{\beta}_{nm}^2. \quad (6)\end{aligned}$$

Коэффициенты разложения в таком случае рассчитываются как

$$\hat{\beta}_{nm} = \iint_D s^{(p)}(x, y) \hat{F}_{nm}(x, y) \hat{\mu}(x, y) dx dy. \quad (7)$$

Поскольку рассматриваемые семейства (5) есть произведение одномерных функций, представленных в табл. 1, то перенос и масштабирование по каждой из осей можно рассматривать отдельно.

Коэффициенты переноса и масштаба для полиномов Лежандра и Чебышева, определённые на конечном интервале, находятся однозначно. В отличие от них, для функций Лагерра и Эрмита в общем случае можно задать лишь коэффициенты переноса, но и они могут быть выбраны почти произвольно. Значения коэффициентов масштаба и переноса приведены в табл. 2.

Поскольку для функций Лагерра и Эрмита невозможно заранее определить оптимальное значение коэффициента масштаба, коэффициенты разложения (7) в общем виде являются функциями от неизвестных коэффициентов α_x и α_y :

$$\begin{aligned}\hat{\beta}_{nm} &= \hat{\beta}_{nm}(\alpha_x, \alpha_y) = \\ &= \frac{1}{\sqrt{\alpha_x \alpha_y}} \iint_D s^{(p)}(x, y) F_{nm}\left(\frac{x - \beta_x}{\alpha_x}, \frac{y - \beta_y}{\alpha_y}\right) \mu\left(\frac{x - \beta_x}{\alpha_x}, \frac{y - \beta_y}{\alpha_y}\right) dx dy. \quad (8)\end{aligned}$$

Таблица 2

Коэффициенты переноса и масштаба		
Вид	Коэффициенты переноса	Коэффициенты масштаба
Полиномы Лежандра, Чебышева	$\beta_x = \frac{1}{2}(x_{\max} + x_{\min}),$ $\beta_y = \frac{1}{2}(y_{\max} + y_{\min})$	$\alpha_x = \frac{1}{2}(x_{\max} - x_{\min}),$ $\alpha_y = \frac{1}{2}(y_{\max} - y_{\min})$
Функции Лагерра	$\beta_x = x_{\min},$ $\beta_y = y_{\min}$	$\alpha_x = \frac{1}{8}(x_{\max} - x_{\min}),$ $\alpha_y = \frac{1}{8}(y_{\max} - y_{\min})$
Функции Эрмита	$\beta_x = \frac{1}{2}(x_{\max} + x_{\min}),$ $\beta_y = \frac{1}{2}(y_{\max} + y_{\min})$	$\alpha_x = \frac{1}{8}(x_{\max} - x_{\min}),$ $\alpha_y = \frac{1}{8}(y_{\max} - y_{\min})$

Представим интеграл (8) по всей области D в виде суммы интегралов по каждой прямоугольной ячейке:

$$\hat{\beta}_{nm} = \frac{1}{\sqrt{\alpha_x \alpha_y}} \sum_{i=0}^{m_x-1} \sum_{j=0}^{m_y-1} \int_{x_i}^{x_{i+1}} \int_{y_j}^{y_{j+1}} \sum_{k=0}^p \sum_{l=0}^p a_{ij}^{(kl)} x^k y^l F_{nm} \left(\frac{x - \beta_x}{\alpha_x}, \frac{y - \beta_y}{\alpha_y} \right) \times \\ \times \mu \left(\frac{x - \beta_x}{\alpha_x}, \frac{y - \beta_y}{\alpha_y} \right) dx dy. \quad (9)$$

Для упрощения аналитических выкладок проведем следующую замену переменных:

$$x_D = \frac{x - \beta_x}{\alpha_x}, \quad y_D = \frac{y - \beta_y}{\alpha_y}. \quad (10)$$

После замены переменных выражение (9) для определения коэффициентов примет вид

$$\hat{\beta}_{nm} = \sqrt{\alpha_x \alpha_y} \sum_{i=0}^{m_x-1} \sum_{j=0}^{m_y-1} \int_{x_{D,i}}^{x_{D,i+1}} \int_{y_{D,j}}^{y_{D,j+1}} \sum_{k=0}^p \sum_{l=0}^p a_{ij}^{(kl)} \left[(\alpha_x x_D + \beta_x)^k (\alpha_y y_D + \beta_y)^l \times \right. \\ \left. \times F_{nm}(x_D, y_D) \mu(x_D, y_D) \right] dx_D dy_D. \quad (11)$$

Перейдём к раздельной записи интегралов:

$$\hat{\beta}_{nm} = \sum_{i=0}^{m_x-1} \sum_{j=0}^{m_y-1} \sum_{k=0}^p \sum_{l=0}^p a_{D,ij}^{(kl)} \int_{x_{D,i}}^{x_{D,i+1}} x_D^k P_n(x_D) \mu_x(x_D) dx_D \times \\ \times \int_{y_{D,j}}^{y_{D,j+1}} y_D^l Q_m(y_D) \mu_y(y_D) dy_D, \quad (12)$$

где

$$a_{D,ij}^{(kl)} = \sqrt{\alpha_x \alpha_y} \sum_{q=k}^p \sum_{r=l}^p a_{ij}^{(kl)} C_q^k \alpha_x^k \beta_x^{q-k} C_r^l \alpha_y^l \beta_y^{r-l}. \quad (13)$$

Представим интегралы в формуле (12) как некоторую функцию от левой и правой границ:

$$I_x^{(kn)}(a, b) = \int_a^b x^k P_n(x) \mu_x(x) dx, \quad I_y^{(lm)}(c, d) = \int_c^d y^l Q_m(y) \mu_y(y) dy. \quad (14)$$

Тогда формулу (12) можно записать в следующем виде:

$$\hat{\beta}_{nm} = \sum_{i=0}^{m_x-1} \sum_{j=0}^{m_y-1} \sum_{k=0}^p \sum_{l=0}^p a_{D,ij}^{(kl)} I_x^{(kn)}(x_{D,i}, x_{D,i+1}) I_y^{(lm)}(y_{D,j}, y_{D,j+1}).$$

Для полиномов Лежандра —

$$I_{\text{Leg}}^{(kn)}(a, b) = \int_a^b x^k \left(\sum_{s=0}^{[n/2]} a_s x^{n-2s} \right) dx = \\ = \sum_{s=0}^{[n/2]} \frac{a_s}{n-2s+k+1} \left(b^{n-2s+k+1} - a^{n-2s+k+1} \right).$$

Для остальных видов функций из табл. 1 интегралы вычисляются аналогичным образом.

В зависимости от вида выбранного семейства ортонормированных функций значения интегралов (14) подставляются в формулу (12) для вычисления коэффициентов разложения. Коэффициенты $\hat{\beta}_{nm}$, рассчитанные для выбранных значений α_x , α_y , β_x и β_y , следует использовать в откорректированной формуле (3) разложения аппроксимируемой функции в ортогональный ряд:

$$\hat{f}(x, y) = \sum_{n=0}^N \sum_{m=0}^{N-n} \hat{\beta}_{nm} \hat{F}_{nm}(x, y) = \sum_{n=0}^N \sum_{m=0}^{N-n} \frac{\hat{\beta}_{nm}}{\sqrt{\alpha_x \alpha_y}} P_n \left(\frac{x - \beta_x}{\alpha_x} \right) Q_m \left(\frac{y - \beta_y}{\alpha_y} \right).$$

Погрешность (6) восстановления исходной функции в большой степени зависит от выбора коэффициентов масштаба. Для семейств ортогональных функций Лагерра и Эрмита оптимизация значений коэффициентов проводится через поиск минимума функции погрешности:

$$- \sum_{n=0}^N \sum_{m=0}^M \left(\sum_{i=0}^{m_x-1} \sum_{j=0}^{m_y-1} \sum_{k=0}^p \sum_{l=0}^p a_{D,ij}^{(kl)} I_x^{(kn)} \left(\frac{x_i - \beta_x}{\alpha_x}, \frac{x_{i+1} - \beta_x}{\alpha_x} \right) \times \right. \\ \left. \times I_y^{(lm)} \left(\frac{y_j - \beta_y}{\alpha_y}, \frac{y_{j+1} - \beta_y}{\alpha_y} \right) \right)^2 \rightarrow \min.$$

Начальное приближение коэффициентов α_x и α_y задается по данным, представленным в табл. 2, а затем осуществляется поиск минимума функции двух переменных по методу Ньютона.

Приведённые результаты легли в основу разработки программного комплекса, используемого для восстановления двумерных плотностей вероятностей по выборкам в ортогональных базисах. Имитационное моделирование, проведенное для апробации полученных результатов, показало, что среди базисов на отрезке $[-1; 1]$ полиномы Чебышева первого рода обеспечивают наилучшую сходимость. Семейства функций Лагерра и Эрмита также показывают хороший результат восстановления для некоторых видов плотностей (Лагерра — для плотностей, близких к экспоненциальному закону, а Эрмита — близких к нормальному закону распределения). В общем случае разложение неизвестной плотности в ортогональный ряд обеспечивает лучшее приближение плотности вероятностей, чем построение по гистограмме билинейного или бикубического сплайна. При этом ортогональный ряд не является кусочно

склеенным и требует хранения меньшего количества рассчитанных коэффициентов. А в отличие от параметрических моделей плотности ортогональная модель инварианта к виду аппроксимируемой характеристики.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Муха В. С. Анализ многомерных данных: Монография. — Мн.: Технопринт, 2004. — 368 с.
2. Суетин П. К. Ортогональные многочлены по двум переменным. — М.: Наука, 1988. — 384 с.
3. Физтенгольц Г. М. Курс дифференциального и интегрального исчисления: В 3-х томах. — М.: Физматлит, 2006. — Т. 2. — 864 с.

Поступила в редакцию 24/X/2008;
в окончательном варианте — 13/II/2009.

MSC: 93B40, 42C05, 33C45

APPROXIMATION OF DOUBLE-DIMENSIONAL DENSITIES OF PROBABILITIES BY ORTHOGONAL BASES

I. A. Lyozin

S. P. Korolyov Samara State Aerospace University,
34, Moskovskoye sh., Samara, 443086.

E-mail: chuchyck@yandex.ru

Approach to approximation of double-dimensional densities of probabilities in rectangle areas by the families of double-dimensional orthogonal functions created by multiplication of single-dimensional orthogonal functions is studied. Calculation formulas of approximation coefficients are quoted at different bases by using primary spline approximation of input data, as well as the meaning of methodical deviation at selected values of scale and transmission coefficients.

Key words: *approximation, orthogonal basis, double-dimensional function, approximation coefficient, density of probabilities, double-dimensional spline.*

Original article submitted 24/X/2008;
revision submitted 13/II/2009.

Lyozin Iliya Alexandrovich, Assist., Dept. of Information System and Technology.